Forest Health Protection

20-05 *OCT, 2020*

Multi-Scale Validation of Aerial Detection Survey Accuracy and Post-Processing Techniques to Depict Mountain Pine Beetle Outbreak Severity and Extent in the Northern Region

Joel M. Egan[1], Jeffrey Kaiden[1], and Jordan Lestina[2]

[1] State and Private Forestry, Forest Health Protection, Missoula, Montana
[2] KSLK Consulting Inc., Missoula, Montana

Figure 1: Photograph showing widespread mountain pine beetle-caused mortality in susceptible lodgepole pine hosts within the Pioneer Mountains on the Beaverhead-Deerlodge National Forest. Photo by Scott Sontag.

United States Department of Agriculture	Forest Service	Northern Region	Region One 26 Fort Missoula Rd Missoula, MT 59804-7203	

Executive Summary

Aerial detection surveys (ADS) currently provide best-available data utilized by Forest Health Protection and Partners to monitor forest pest status and trends across the United States. Recently, data collection protocols changed and the metric collected to capture severity of tree mortality transitioned from trees/acre to percent mortality in 2016. Post-processing techniques for ADS data were developed to 1) reduce biases inherent within ADS data to improve quality and 2) crosswalk data collected with historic and modern metrics to facilitate comparisons and cumulative analyses.

This report assesses three respective post-processing techniques and the accuracy of aerial detection survey data through spatially explicit comparisons with an independently derived validation dataset. Cumulative ADS data for mountain pine beetle (*Dendroctonus ponderosae* Hopkins, MPB) from 1999-2015 were validated with a mortality dataset within Region 1's Existing Vegetation Map (R1-VMap). Comparisons occurred across two study areas, spanning over 11 million acres, across the Beaverhead-Deerlodge and Helena-Lewis and Clark National Forests in central Montana. Within study areas, analyses were conducted at three spatial scales: 1) localized polygon-level, 2) base-level (HUC12 subwatersheds), and 3) mid-level (National Forest).

At the polygon-level, 72% of the 11.2 million study areas surveyed had ***area mapped*** with ADS validated to correctly represent MPB-caused damage while 22%, or 2.4 million acres, had ADS area mapped where no damage occurred. False positives were adjacent to validated damage, indicating ADS polygons were mapped in accurate locations with spatial extent overestimation bias detected. Severity class validation, within three percent mortality classes, occurred for each post-processing method for those locations where ADS and R1-VMap datasets both indicated presence of MPB-caused damage. ADS severity estimates were accurate across a low percentage (range 29-47%) of the subset area depending on study location and post-processing method.

Base-level validation occurred after ***area mapped*** totals were scaled to ***damage area*** to remove live tree areas from polygons based on respective post-processing methods with data summed to subwatershed (HUC-12) unit for analysis. Across all methods, a majority of subwatersheds had high rates of correct classification that were within ± 1500 acres of reference data. The $FHP_{R1\&R4}$ method had greatest percent of correctly classified subwatersheds (82-83%) while Histogram Matching and Meddens methods had reduced accuracy (67-73%). Regression analysis identified significant, positive relationship between ADS ($FHP_{R1\&R4}$ method) and reference ***damage area*** with 77-83% of variation between variables explained.

At mid-level (National Forest), ***area mapped*** was inaccurate and nearly double (1.2-1.4 million acres) the area of total pine habitat identified in reference data. This resulted from liberal spatial extent attribution identified at the polygon-level. Post-processing conversions from ***area mapped*** to ***damage area*** reduced live tree areas from polygons, while adjusting for spatial extent overestimation bias. Post-processing methods performed differently to reduce this bias as the Histogram-matching and Meddens method's mid-level ***damage area*** totals exceeded reference data by 200% and 300% while $FHP_{R1\&R4}$ method had ***damage area*** totals within a reasonable 15% and 45% of reference data across respective study sites.

ADS classification accuracy was improved as scale of inference was increased from polygon to base- and mid-levels. Low severity class accuracy rates occurred within individual polygons; however, accuracy improved substantially after post-processed data was summed to coarser scales as ADS severity attribution errors of over- and underestimation balanced out. Further, the $FHP_{R1\&R4}$ post-processing technique was validated as most-accurate at both base- and mid-levels as its conservative approach was sufficient to adjust for the spatial extent overestimations identified at polygon-level.

Our findings indicate post-processed ADS data had high levels of accuracy in representing multi-year, bark beetle impacts across our study areas when post-processed with bias-reduction techniques. However, we document substantial issues in reporting ***area mapped*** (also called 'acres with mortality' or 'raw acres') totals without post-processing. Our study and others suggest reporting of ***area mapped*** statistics should include caveats that this statistic incorporates 1) spatial extent overestimation bias and 2) various proportions of damaged and non-damaged (live-tree areas) within mapped polygons. Post-processing is recommended to derive ***damage area*** report pest status along with ***area mapped*** totals and to utilize ***damage area*** as a standardized statistic for trend analyses and pest impact comparisons.

Introduction

Aerial detection surveys (ADS), collected by U.S. Forest Service, Forest Health Protection programs and Partner agencies are currently the best available source of data for reporting forest damage and pest trends across vast landscape scales (McConnell et al. 2000). Aerial surveys are conducted annually by observers in aircraft trained to visually identify host species and damage agents by relying on foliage change following the year of attack. Observed pest impacts are sketched as points and polygons on an electronic map and attributes such as tree host, damage agent, and severity of damage are recorded. Certified surveyors conduct ADS missions to collect data in nationally standardized format then post-processing converts data to useable information to report-out to resource managers and the public (Man 2012). ADS data is utilized to publish insect/disease monitoring reports and by local resource managers to inform forest management planning and implementation.

While aerial surveys are an economic and efficient method to monitor forest disturbance conditions, these data have limitations. Data are collected through observational means rather than direct measurements. Aerial observations tend to vary between surveyors due to differences in skill, sketching preferences (i.e. lumping vs. splitting), viewing conditions, and surveyor tolerance of physical discomfort associated with moving at high speeds in small, fixed-wing aircraft. Therefore, errors in estimations of spatial extent, spatial location, severity, and damage agent attribution are potential sources of bias when aggregating yearly damage data into multi-year, cumulative summaries (see Egan et al. 2019 for bias discussion).

To-date, few assessments have validated ADS survey data quality by comparing ADS to 'truth' datasets derived from ground surveys or remotely sensed data. Coleman et al. (2018) recently published validation work that indicated surveyor findings are repeatable and damage agents identified were accurate to genus, and sometimes species, taxonomic levels. Johnson and Ross (2004) found that ADS data summaries were reasonable to consider at coarse spatial scales but are suspect at fine-scales after assessing the precision of ADS mapping locations. Multiple studies have also reported underestimation bias within ADS data severity estimates as well as overestimation of spatial extent for area impacted on the ground (Meddens et al. 2012; Backsen and Howell 2013; Coleman et al. 2018).

These studies led to development of post-processing methods to reduce bias within ADS data (see Egan et al. 2019). In short, techniques are aimed to reduce underestimation bias and increase accuracy by categorizing ADS data into a 3-class percent mortality system based on severity data collected as either trees per acre estimates or percent mortality within 5-class categories. This method was validated as promoting substantially increased accuracy rates using 3-classes across two study areas in which ADS data was compared to Worldview II high-resolution remotely sensed imagery (Bright et al. 2020).

Along with percent mortality classifications, various post-processing techniques have been developed to cross-walk severity data collected in surveys between historic survey methods (severity collected with trees per acre variable) to modern methods (severity collected with percent mortality class variable).

These alternate post-processing methods include the $FHP_{R1\&R4}$ method (Egan et al. 2019), the histogram matching method (Zweifler 2019), and the Meddens method (Meddens et al. 2012). Hicke et al. (2020) compared each of these alternate methods with ADS data from 1998-2018 and evaluated cross-walk results to calculate annual and total beetle-caused tree mortality totals across the western U.S. Conclusions suggested the $FHP_{R1\&R4}$ technique to be most reasonable, but advised further validation comparing results to an independent disturbance dataset. This report addresses that validation need.

In this report, we aimed to determine: 1) ADS data quality and validity for use at various scales of inference and 2) utility of various post-processing methods used to cross-walk trees per acre to percent mortality estimates and ***damage area*** calculations. Specifically, our analysis compared multi-year ADS data from 1999-2015 for mountain pine beetle (*Dendroctonus ponderosae* Hopkins, MPB)-caused mortality to an independent mortality dataset produced by Region 1's Vegetation Map (R1-VMap). Comparisons occurred across two extensive study areas in the Beaverhead-Deerlodge and Helena-Lewis and Clark National Forests of central Montana. Analyses were conducted at three different spatial scales at mid-level (i.e. National Forest), base-level (i.e. subwatersheds), and at a localized, polygon-level (Nelson et al. 2015).

Methods

Study areas

Study locations included forested areas with pine species present across the Helena-Lewis and Clark and Beaverhead-Deerlodge National Forests that spanned 5.7 and 5.4 million acres, respectively. Forests are located in central Montana, primarily on the eastern slope of the northern Rocky Mountains.

Geospatial Datasets

Existing Vegetation Database (R1-VMap) Data

The R1-VMap database (U.S. Department of Agriculture Forest Service 2018) has been published by the Northern Region Geospatial Group since 2000, and is designed to be used for Forest and Landscape level analysis that follows the standards set forth in the *Existing Vegetation Classification and Mapping Technical Guide* (Brohman and Bryant 2005) and the *Region 1 Multi-level Classification, Mapping, Inventory, and Analysis System* (Berglund et al. 2009). It is a spatially explicit, thematic, polygon-based product derived from ground plots, remotely sensed, topographic, and climatic data modeled to provide information about the extent, composition, and structure of forest vegetation. R1-VMap is a multi-level, existing vegetation geospatial database used to produce four primary map products: lifeform, tree canopy cover class, tree diameter, and tree dominance. Additional add-ons to the database include dead tree canopy which were created for both study areas used in this analysis (Ahl et al. 2018).

The National Agriculture Imagery Program (NAIP) digital orthophotography and field data for the Helena – Lewis & Clark (H-LC) National Forest VMap update data were gathered in 2011 and was published in 2014 (U.S. Department of Agriculture Farm Service Agency 2011). The Beaverhead–Deerlodge (BVRD) National Forest VMap update was published in 2018 with the field data collected in 2017 across 4,000 reference plots. Field data collection is summarized in this story map, (Kaiden 2018). NAIP imagery used for the BVRD originated in 2013, but 2016 and 2017 imagery could not be used due to smoke. Thus, lower-resolution Sentinel 2 (European Space Agency 2015) satellite data from July 18, 2016 to August 8, 2017 was used to supplement the NAIP imagery and create the final R1-VMap product for BVRD.

HUC12 Subwatersheds Data

The Watershed Boundary Dataset (WBD) maps the full areal extent of surface water drainage for the U.S. using a hierarchical system of nesting hydrologic units at various scales, each with an assigned hydrologic unit code (HUC; USGS and USDA NRCS 2013). In the hydrologic unit code hierarchy, the HUC-12 or 6^{th}-level, is a subwatershed that captures tributary systems within a local drainage. There are approximately 160,000 HUC-12 units across the U.S. that average 40 miles^2 in area (range 10,000-40,000 acres).

Fire Perimeter Data

The Northern Region's fire perimeter geospatial dataset incorporates wildland fire perimeters for all incidents 10 acres or greater in size and some smaller events may be included as well. Data are maintained at the Forest/District level to track the area affected by wildland fire with a target scale of 1:24,000. Each of these fire perimeter records have a corresponding record in FIRESTAT, which is the authoritative data source for all wildland fire occurrences on National Forest System lands (USDA FS 2016). Polygons include all burned areas as well as unburned islands, rocky outcrops, water bodies, and marshlands within the fire perimeter.

Aerial Detection Survey Data

Aerial Detection Survey (ADS) data are collected annually by FHP and Partners. Data are collected through methodologies described within McConnell et al. (2000) for data collected from 1999-2015. Primary variables collected used in analysis included where damage agent was attributed as MPB for pine host trees, damage severity estimated with trees per acre (TPA) mortality rates, and spatial information attributed through mapping that included location, spatial extent, and no damage (as indicated through non-attribution within surveyed areas).

ADS data were spatially aggregated from 1999-2015 annual surveys to represent cumulative damage caused by the MPB outbreak (Lestina et al. 2019). For this, annual polygon data were temporally combined into a cumulative data layer with TPA values summed where spatial overlap occurred. Survey coverage data layers were aggregated in similar manner to depict cumulative years of survey coverage. From these data, any portions of the study area that did not have 8 years or greater of survey coverage were excluded from the analysis.

MPB-caused damage severity information was attributed as the rate of mortality per given area (TPA) during the 1999-2015 outbreak period. This information is collected in tandem with variably sized spatial points or polygons with survey methods that confound spatial extent and severity. As such, 1) spatial extents incorporate both live and dead trees within area quantifications and 2) spatial extent and severity are interdependent and cannot be considered separately. Thus, post-processing methodologies described subsequently consider spatial extent and severity simultaneously to represent damage acre totals (hereafter referred to as ***damage area*** which is synonymous with the *severity-weighted area* technical terminology used in our Egan et al. 2019 methods paper). Additionally, post-processing methods converted ADS data from TPA to percent mortality metrics to match percent mortality metric modeled within the R1-VMap reference dataset.

Aerial Survey Data Post-Processing: Trees per Acre to Percent Mortality and Severity-Weighting

Aerial Survey Data Post-Processing: FHP R1&R4 method

We calculated the ***damage area*** statistic that represents aggregate, high-severity (>30%) impact area total using methods described in Egan et al. (2019). Processing was sequential and included: 1) aggregating annual 1999-2015 ADS and summing overlapping spatial data trees per acre (TPA) values to yield cumulative mortality (see Lestina et al. 2019); 2) converting cumulative mortality trees per acre values to broad-category classifications based on percent mortality ranges for low (<10%, based on <10 TPA), moderate (10-30%, based on 10-30 TPA), and high (30-100%, based on >30 TPA); then 3) using the relative midpoints from low and moderate classes to scale area totals within polygons to a high-severity baseline for ***damage area*** calculations. Scaling factors derived from midpoint relativity were 0.0769 for low-severity damage, 0.3077 for moderate-severity, and 1.0 (unscaled) for the high-severity damage baseline. Then, post-scaled area totals were summed to yield ***damage area*** totals used for analysis.

Aerial Survey Data Post-Processing: Histogram Matching method

The histogram-matching method documented in Zweifler (2019) was built to support data assessments using the legacy TPA metric and was reversed to convert ADS TPA variables to percent mortality. For this, data are converted based on aligning histogram distributions of multi-year data with data collected using respective TPA and percent mortality metrics. ADS data used for histogram matching were from 2011-2015 (TPA metric collected) and 2016-2017 (5-class percent mortality data collected). Aerial survey cumulative TPA data from 1999-2015 were converted final ***damage area*** products by: 1) collapsing percent mortality classes from 5- to 3-classes and classifying ADS TPA values into respective class of low (<10%, based on <5 TPA), moderate (10-30%, based on 5-9.9 TPA), or high (>30%, based on $\geq$ 10 TPA); 2) assigning TPA values into representative percent mortality classes; and 3) using the relative midpoints from low and moderate classes (same methodology as described above for the $FHP_{R1\&R4}$ method) to scale area totals within polygons to a high-severity baseline. Then, post-scaled area totals were summed to yield ***damage area*** totals used for analysis.

Aerial Survey Data Post-Processing: Meddens method

The Meddens method was derived after substantial underestimation bias was identified in ADS data following comparisons to remotely sensed data as documented in Meddens et al. (2012). In short, this method calculates the ratio of impacted pine crown area relative to the total polygon area and yields an estimate of continuous percent mortality. For this, total cumulative TPA data from ADS, within respective polygons, were multiplied by 1) crown area scaling factor (representing the average crown area each respective pine tree occupies within an acre); 2) the underestimation bias scaling factor of 13.6; and 3) 100 to convert data from proportion to percent scales. The equation for this conversion is: percent mortality = TPA × CROWN × 100 × 13.6. Note, this conversion led to numerous data points with >100% mortality which were classified as 100%.

Then, continuous percent mortality data from Meddens method were 1) assigned to low (<10%), moderate (10-30%), and high (>30%) severity classes; 2) ***damage area*** totals were scaled to an aggregate, high-severity baseline based on severity class midpoint relativity (as described above for $FHP_{R1\&R4}$ method); and 3) summed to yield ***damage area*** totals used for analysis.

Aerial Survey Data Post-Processing: TPA to % Mortality Conversion Thresholds

Post-processing methods had variation in conversion factors relating TPA assignment into low, moderate, and high percent mortality classes (Table 1). Thresholds for classification were relatively more conservative for the $FHP_{R1\&R4}$ method relative to the Histogram matching and Meddens methods.

Table 1: Table depicts ranges of trees per acre (TPA) values used to classify Aerial Detection Survey damage into percent severity classes based on respective post-processing methods.

ADS Post-Processing Method	Species	Crown Area Factor	Low (< 10%)	Moderate (10-30%)	High (>30%)
Meddens	lodgepole pine	0.002144875	< 3 TPA	3-10.9 TPA	>= 11 TPA
Meddens	ponderosa pine	0.004146429	1 TPA	2-5 TPA	>= 6 TPA
Meddens	whitebark/limber*	0.00308264	< 3 TPA	3-7.9 TPA	>= 8 TPA
Histogram Matching	All	-	< 5 TPA	5-9.9 TPA	>= 10 TPA
$FHP_{R1\&R4}$	All	-	< 10 TPA	10-30 TPA	> 30 TPA

**** Note: ADS host species in R1/R4 for limber and whitebark pines are averaged as ADS host data combines both hosts into a five-needled pine category***

Analysis Methods

Study Area Delineation

Study areas were delineated across the H-LC and BVRD NFs using the full extent of all 6th–level (10-digit) watersheds that intersect NFS ownership (range 40,000-250,000 acres). The prior NF included primarily National Forest System-managed lands only while the latter incorporated adjacent private, state, and BLM-managed lands. The northeast boundary of the BVRD is delineated by the H-LC and the southern boundary of the BVRD follows the boundary of the USFS Northern Region. The final boundary was determined aligning with HUC12 boundaries, essentially, following the course of the Clark Fork River.

Geospatial Processing Methods

R1-VMap data for the H-LC and BVRD NFs delineate the boundaries for this study; thus, data from ADS, fire perimeter, and the National Hydrography dataset were clipped to their extents. Then, the R1-VMap data was filtered to include only those areas that incorporated tree species vulnerable to the mountain pine beetle: ponderosa pine, lodgepole pine, whitebark pine, and limber pine in locations with at least 40% of total basal area having pine species presence. Locations with wildland fire incident occurrence were excluded from this study. Similarly, only those areas that had ADS survey coverage $\geq$ 50% (8 years) of the 1999-2015 evaluation period were included.

The R1-VMap, ADS damage for MPB, and HUC12 datasets were merged and area calculations summed for total ***damage area*** values for MPB-caused mortality estimated by R1-VMap as well as the ADS data following post-processing with the Meddens, Histogram Matching, and $FHP_{R1\&R4}$ methods, respectively. All acreage and severity weighted acre values were calculated with an equal-area projection (Albers Equal Area Conic).

Analysis method

Analysis occurred across each respective Forest study area at three spatial scales to evaluate different sources of potential error within ADS data collected for MPB-caused mortality from 1999-2015. Specifically, four potential sources of error and bias within ADS data collected were evaluated through reference data comparisons to quantify 1) spatial location errors; 2) mapping extent errors; 3) severity estimation errors; and 4) other, unknown errors ingrained within post-processing methodologies.

First, a fine-scale spatial overlap analysis occurred at the polygon-level to assess spatial location, mapping extent, and severity estimation errors within the ADS dataset. The ADS data were compared to R1-VMap reference data to identify 1) ADS omission/commission errors and spatial agreement with validation data and 2) where spatial agreement occurred, validation of ADS severity estimates for each of the post-processing methods.

Then, a base-level analysis used a subwatershed unit of replication and total ***damage area*** to 1) descriptively assess the orientation and magnitude of deviation between ADS and R1-VMap ***damage area*** totals for each post-processing method and 2) inferential analysis with linear regression modeling to determine relationship significance and orientation between the ADS estimates and R1-VMap (SAS v9.1).

Finally, a mid-level analysis compared 1) the extent of ***areas mapped*** by ADS as impacted by MPB with R1-VMap reference data; and 2) ***damage area*** (representing aggregate MPB damage after excluding proportion of mapped area containing live trees) from ADS data modified with respective post-processing methods to R1-VMap reference ***damage area***.

Results

Polygon-level Validation of Damage Locations and Severity Estimates from Aerial Surveys

At the local polygon-level, which is the unit of attribution for both ADS and R1-VMap data, the spatial locations of ADS-mapped MPB-caused tree mortality were intersected with R1-VMap reference data to assess spatial location attribution errors (Figures 2 and 3, left panel maps; Table 2). This spatial location error assessment only considered the presence/absence of ADS ***area mapped***; thus, there is no post-processing methods comparison for this analysis. Overall, presence and absence of MPB-caused tree mortality was mapped with ADS correctly over 72% of the combined study areas (Table 2).

ADS damage locations differed from reference data in some areas as 7% of the combined study area had ADS failure to map MPB damage where it occurred (omission errors) and 22% had MPB damage mapped where none occurred (commission error) (Table 2). Evaluation of commission error occurrences across study area maps (Figures 2 and 3 - left panels) identified these error locations were immediately adjacent to correctly mapped locations of MPB-caused damage. This indicates commission errors were largely the result of liberal mapping of spatial extent (i.e. spillover) that overestimated MPB-caused damages that were otherwise correctly attributed.

Table 2: Total *area mapped* (acres and relative %) of Aerial Detection Survey Spatial Location Validation based on comparisons to R1-VMap Data.

Study Location	Agreement ADS & R1-VMap MPB Presence	Agreement ADS & R1-VMap MPB Absence	Disagreement ADS Omission (false negative)	Disagreement ADS Commission (false positive)	Total Area	Percent Agreement
H-LC NF	610,442	3,506,786	426,942	1,189,444	5,733,614	72%
BVRD NF	844,825	3,035,108	305,661	1,247,015	5,432,609	71%
Combined H-LC & BVRD	1,455,267	6,541,894	732,603	2,436,459	11,166,223	72%

Severity classes (low, moderate, high) validation occurred for a subset of the data intersected for the prior evaluation where ADS and R1-VMap datasets both indicated presence of MPB-caused damage. Severity class attribution was evaluated for ADS data, for each post-processing method, by quantifying total acres and relative subset area that had ADS and reference data agreement on disturbance severity or ADS severity class that over- or under-estimated severity as compared to reference data.

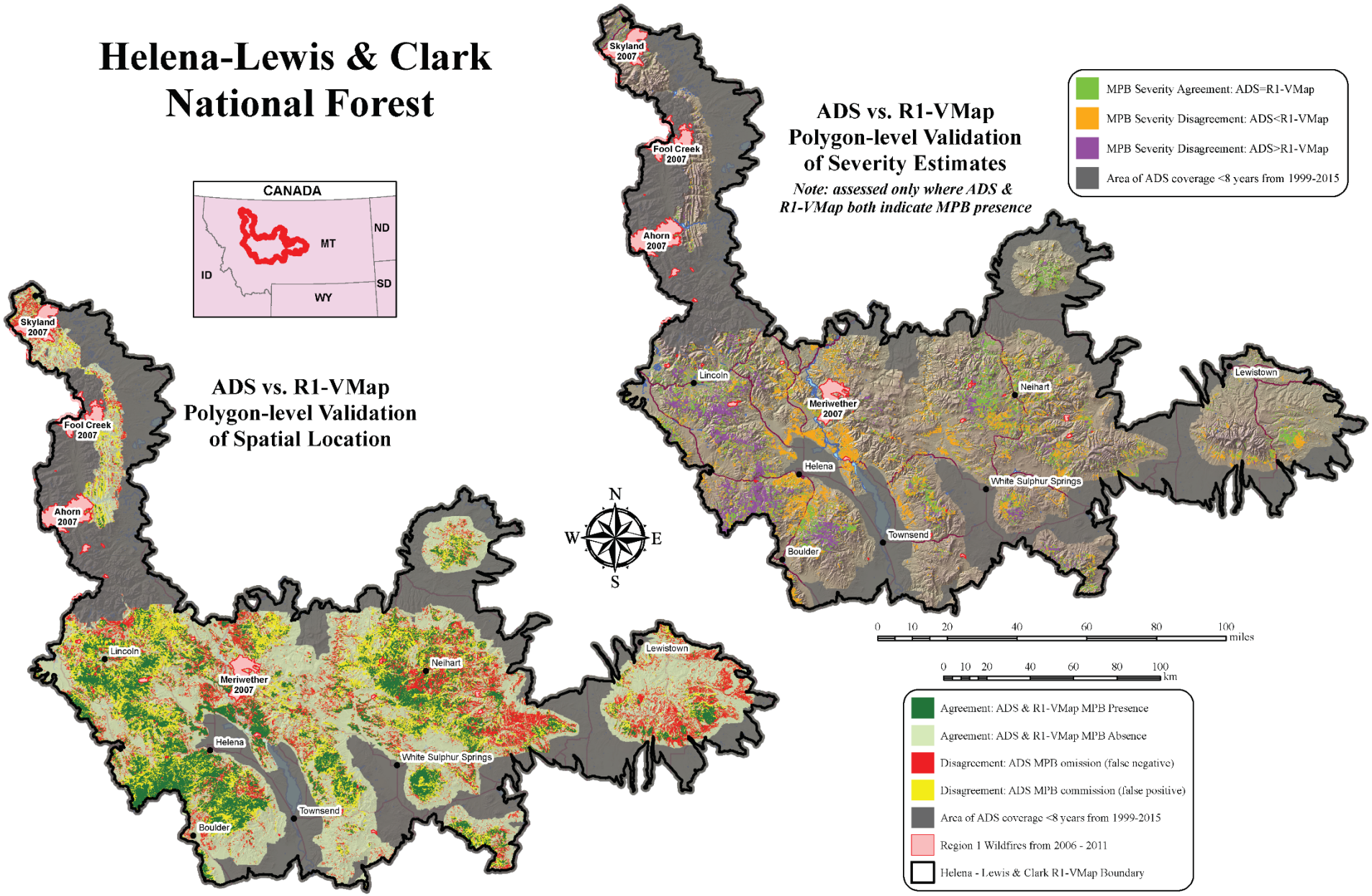

Figure 2: Evaluation of MPB-caused mortality ADS polygon spatial locations and severity estimates using $FHP_{R1\&R4}$ post-processing method in comparison to R1-VMap reference data on the Helena – Lewis and Clark National Forests.

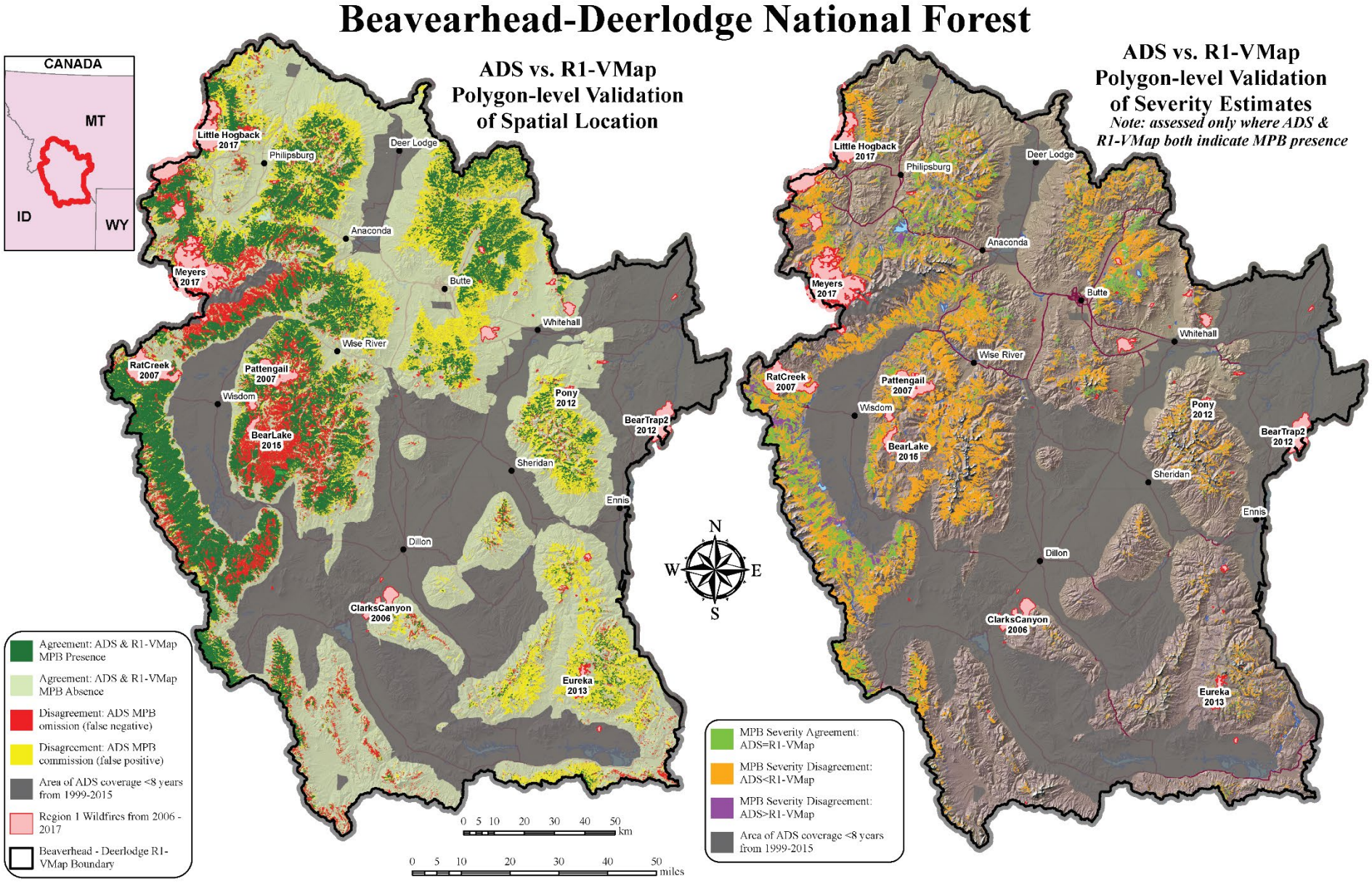

Figure 3: Evaluation of MPB-caused mortality ADS polygon spatial locations and severity estimates using $FHP_{R1\&R4}$ post-processing techniques in comparison to R1-VMap reference data on the Beaverhead – Deerlodge National Forests.

Severity classes estimated by ADS, across all post-processing methods, consistently had <50% of total area correctly classified by ADS estimates based on R1-VMap reference severity data for those locations (Table 3). The $FHP_{R1\&R4}$ method had highest classification accuracy (37%) in the H-LC NF study site while the Meddens method yielded highest accuracy (45%) in the BVRD NF study site. The $FHP_{R1\&R4}$ method tended to underestimate severity across both study area locations (Figures 2 and 3 - right panels only) while the Meddens and Histogram-matching methods tended to overestimate severity.

Table 3: Total area (acres and relative %) of Aerial Detection Survey severity class estimates by validation class based on comparison to R1-VMap reference data.

Study Location	Post-processing Method	Agreement ADS Correct Severity Class	Disagreement ADS Underestimated Severity Class	Disagreement ADS Overestimated Severity Class	Total	Percent Agreement	Percent Underestimated	Percent Overestimated
H-LC NF	$FHP_{R1\&R4}$	225,333	220,036	165,073	610,442	37%	36%	27%
H-LC NF	Histogram Matching	178,673	118,959	312,811	610,442	29%	19%	51%
H-LC NF	Meddens	198,013	81,820	330,608	610,441	32%	13%	54%
BVRD NF	$FHP_{R1\&R4}$	233,442	563,523	47,861	844,825	28%	67%	6%
BVRD NF	Histogram Matching	339,021	238,672	267,132	844,825	40%	28%	32%
BVRD NF	Meddens	380,218	225,397	239,209	844,825	45%	27%	28%
Combined	$FHP_{R1\&R4}$	458,775	783,559	212,934	1,455,268	32%	54%	15%
Combined	Histogram Matching	517,694	357,631	579,942	1,455,268	36%	25%	40%
Combined	Meddens	578,232	307,217	569,817	1,455,265	40%	21%	39%

Base-level validation at subwatershed-scale

The base-level analysis used subwatersheds as spatial unit of replication to validate summed ***damage area*** totals from ADS estimates with R1-VMap reference data. For this, each post-processing method was evaluated for deviation to R1-VMap data with descriptive statistics, scatter plots, and percent classification frequency within broad deviation classes. Then, statistical relationship between ADS data and R1-VMap validation data was evaluated within linear regression solely for the post-processing method that yielded results closest to the R1-VMap validation data.

Descriptive statistics indicate deviation between ADS-estimated and R1-VMap reference ***damage area***, calculated for each respective subwatershed, was lowest with the $FHP_{R1\&R4}$ post-processing method as average ***damage area*** deviation was within a few hundred acres of reference data and median deviation was zero (Table 4). Histogram-matching and Meddens methods had larger ***damage area*** deviations that averaged over a thousand acres across subwatersheds consistently for both study sites (Table 4).

Table 4: *Damage area* differences and over-/under-estimations proportions of ADS data aggregated at the HUC12 subwatershed scale based on comparisons to R1-VMap reference data.

	Damage area difference: ADS minus R1-VMap average [median]			Subwatersheds overestimated Count [% of total]			Subwatersheds underestimated Count [% of total]		
Study Area	**$FHP_{R1\&R4}$**	**Histogram Matching**	**Meddens**	**$FHP_{R1\&R4}$**	**Histogram Matching**	**Meddens**	**$FHP_{R1\&R4}$**	**Histogram Matching**	**Meddens**
H-LC NF	380 [0.0]	1493 [121]	1512 [204]	200 [46%]	294 [68%]	313 [72%]	232 [54%]	138 [32%]	119 [28%]
BVRD NF	-184 [0.0]	1392 [57]	1315 [80]	204 [47%]	319 [75%]	321 [75%]	223 [52%]	108 [25%]	106 [25%]
Combined	92 [0.0]	1412 [59]	1385 [101]	404 [47%]	613 [71%]	634 [73%]	455 [53%]	246 [29%]	225 [26%]

Scatter plots of deviation values (ADS – R1-VMap ***damage area*** within each subwatershed) by post-processing method show the $FHP_{R1\&R4}$ method led to a balanced distribution of under- and over-estimated damage area within subwatersheds of both study sites (Figure 4). Conversely, the Histogram-matching and Meddens methods primarily led to ADS ***damage area*** overestimations. Overall, the percent of subwatersheds that respectively under- vs. over-estimated ***damage area***, across combined study areas, were balanced with $FHP_{R1\&R4}$ method (53% vs. 47%, respectively) while being heavily skewed towards

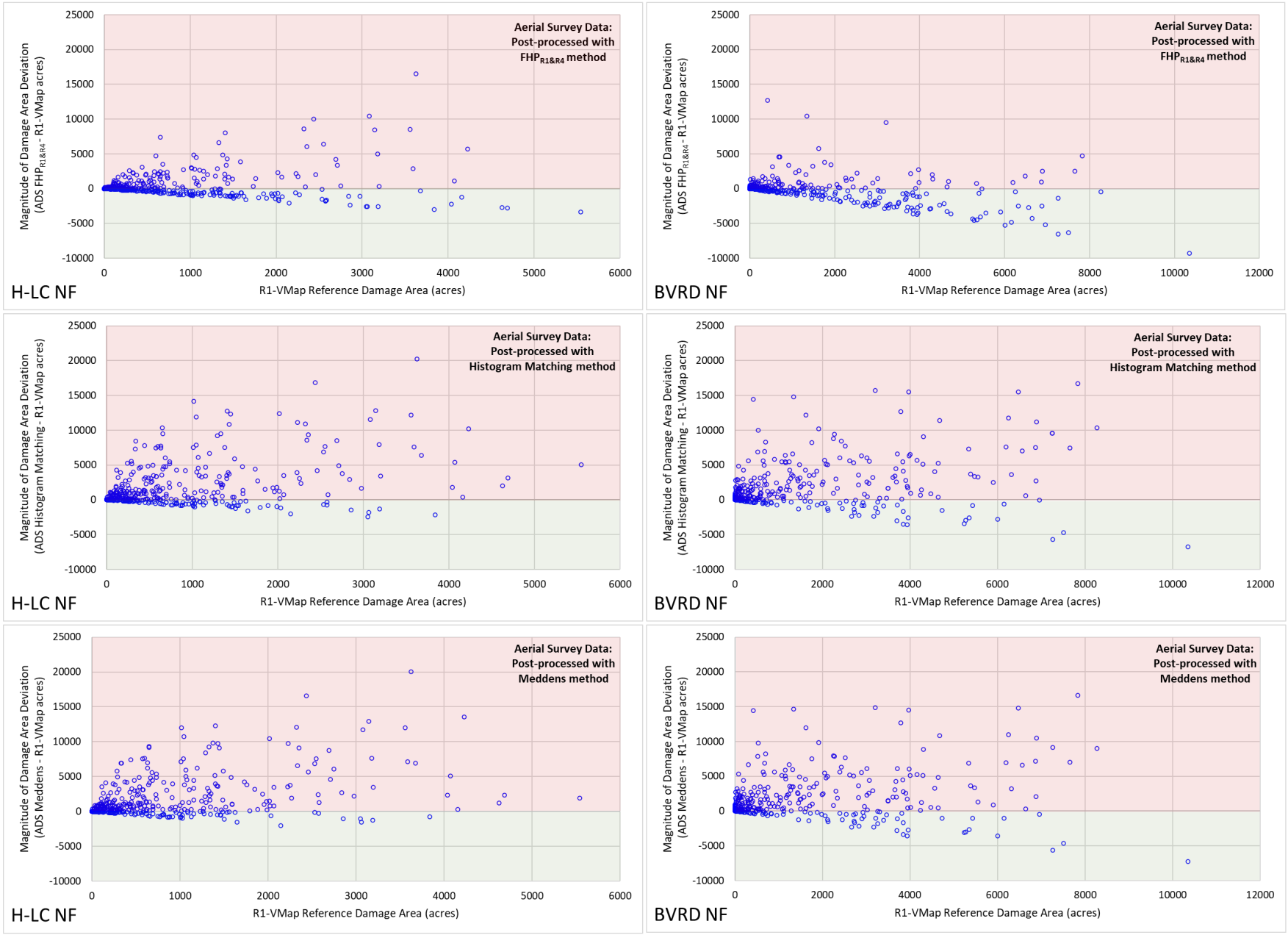
Aerial Survey Data: Post-processed with FHP R1&R4 method
Magnitude of Damage Area Deviation (ADS FHP R1&R4 - R1-VMap acres)
R1-VMap Reference Damage Area (acres)
H-LC NF
BVRD NF
Aerial Survey Data: Post-processed with Histogram Matching method
Magnitude of Damage Area Deviation (ADS Histogram Matching - R1-VMap acres)
Aerial Survey Data: Post-processed with Meddens method
Magnitude of Damage Area Deviation (ADS Meddens - R1-VMap acres)

Figure 4: Scatter plots of deviation values (ADS – R1-VMap *damage area* by subwatershed) for each post-processing method. Aerial survey-estimated *damage area* overestimations are highlighted in red, while underestimations are in green.

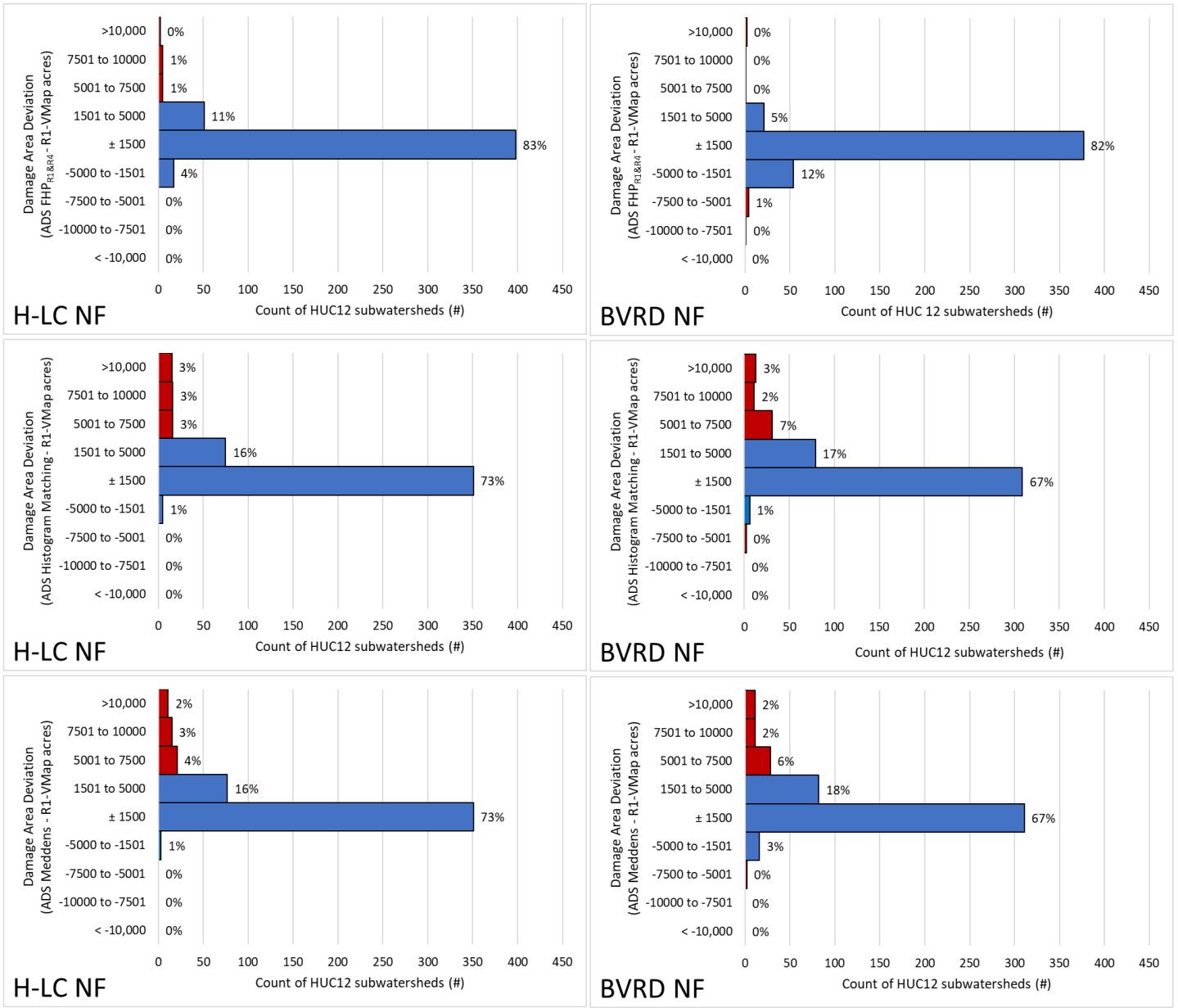
Damage Area Deviation (ADS FHP R1&R4 - R1-VMap acres)
>10,000
7501 to 10000
5001 to 7500
1501 to 5000
± 1500
-5000 to -1501
-7500 to -5001
-10000 to -7501
< -10,000
0% 1% 1% 11% 83% 4% 0% 0% 0%
0 50 100 150 200 250 300 350 400 450
Count of HUC12 subwatersheds (#)
H-LC NF
Damage Area Deviation (ADS FHP R1&R4 - R1-VMap acres)
0% 0% 0% 5% 82% 12% 1% 0% 0%
Count of HUC 12 subwatersheds (#)
BVRD NF
Damage Area Deviation (ADS Histogram Matching - R1-VMap acres)
3% 3% 3% 16% 73% 1% 0% 0% 0%
H-LC NF
Damage Area Deviation (ADS Histogram Matching - R1-VMap acres)
3% 2% 7% 17% 67% 1% 0% 0% 0%
BVRD NF
Damage Area Deviation (ADS Meddens - R1-VMap acres)
2% 3% 4% 16% 73% 1% 0% 0% 0%
H-LC NF
Damage Area Deviation (ADS Meddens - R1-VMap acres)
2% 2% 6% 18% 67% 3% 0% 0% 0%
BVRD NF

Figure 5: Distribution of area deviation between the ADS $FHP_{R1\ \&\ R4}$ 3-Class and R1-VMap 3-Class tree mortality estimates within subwatershed units on the Helena – Lewis & Clark National Forests (left panels) and the Beaverhead Deerlodge National Forests (right panels) for each post-processing method. Blue bars indicated frequency of subwatersheds with ADS deviations from reference data that were within ± 5000 acres and red bars indicate subwatersheds with deviations exceeding ± 5000 acres.

overestimation with the Histogram-matching (29% vs. 71%) and Meddens (26% vs. 73%) methods (Table 4).

Subwatersheds were then classified within classes based on magnitude of deviation for ***damage area*** between ADS-estimates and R1-VMap reference data (Figure 5). Across all post-processing methods, majority of subwatersheds had ADS ***damage area*** totals within ± 1500 acres of R1-VMap reference data. However, the $FHP_{R1\&R4}$ method had the greatest percent of subwatersheds (82 & 83% for BVRD and H-LC NF study locations, respectively) within ± 1500 acres of R1-VMap totals while reduced rate of classification accuracy occurred with the Histogram-matching (67-73%) and Meddens (67-73%) methods (Figure 5).

Regression modeling occurred solely with ADS data processed with the $FHP_{R1\&R4}$ method to verify relationship between ADS data and R1-VMap validation data (Figure 6). Modeling verified a statistically significant, positive relationship between ADS and R1-Vmap ***damage area*** totals within each study site. After transforming variables (log+1) to meet model assumptions of homogeneous error variances for residuals, modeling found that the ADS data represented 77-82% of the variation within the R1-VMap validation data. Thus, the data derived from ADS with $FHP_{R1\&R4}$ post-processing was proportional to R1-VMap ***damage area*** totals in a consistent, measurable manner across the study sites.

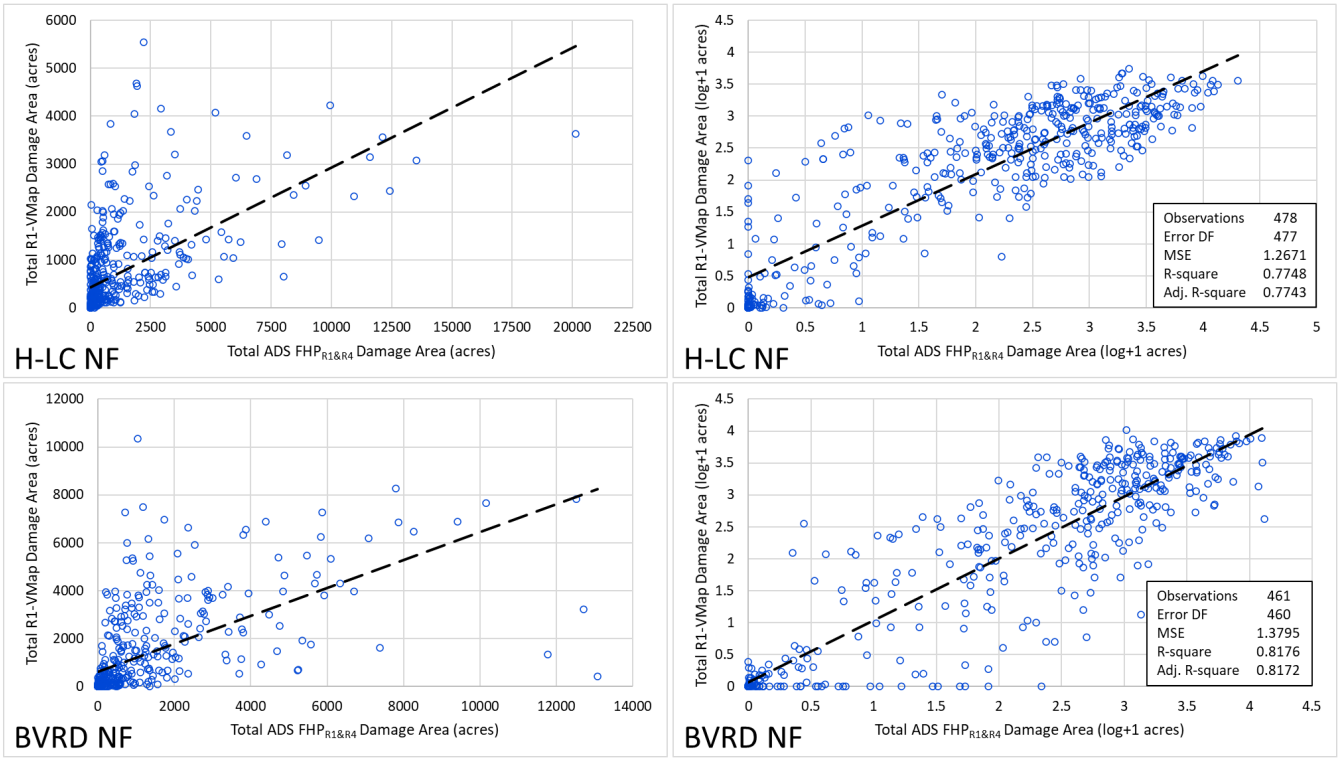

Figure 6: Logistic regression model fits for the ADS $FHP_{R1\&R4}$ 3-Class and R1-VMap 3-Class tree mortality estimations at the subwatershed level on the Helena – Lewis & Clark National Forests (top) and Beaverhead Deerlodge National Forests (bottom). Left panels depict the regression model for only *damage area* estimations and the right panels depict those *damage area* values transformed by log+1.

Mid-Level Validation of Aerial Survey Data at National Forest-scale

Mid-level validation and assessment of post-processing techniques occurred using ***area mapped*** totals from ADS (which includes live and dead tree area) and ***damage area*** statistic which removes live tree area from mapped extent through down-weighting. After removal of areas representing live trees, ***damage area*** as a statistic is representative of aggregate, high-severity damage (>30% mortality = high severity).

Area mapped estimated by ADS, prior to post-processing, yielded nearly double, 82% and 74% higher, or 1.2 and 1.4 million acres of additional ***area mapped*** with MPB activity compared to the R1-VMap validation data (Figure 7). ADS ***area mapped*** totals far exceeded the total pine coverage area by > 1 million acres across respective BVRD and H-LC NF study areas.

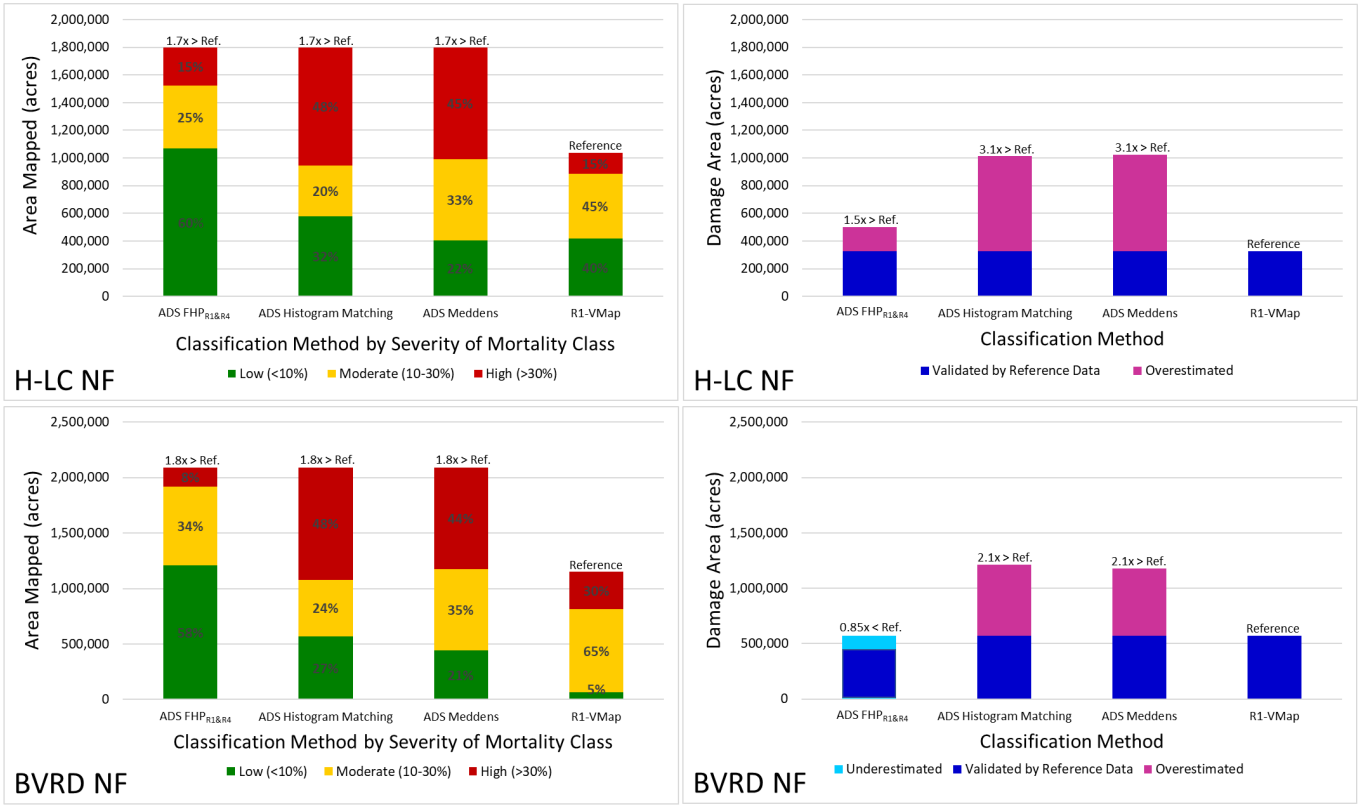

Figure 7: Helena-Lewis & Clark National Forest (top) and Beaverhead - Deerlodge National Forest (bottom) total *area mapped* and *damage area* (acres) MPB-caused tree mortality comparisons calculated from each classification method.

Post-processing methods classified ***area mapped*** into 3-classes of severity differently based on divergent trees per acre to percent mortality cross-walk thresholds (Figure 7 – left bar charts, Figures 9 and 10). The $FHP_{R1\&R4}$ method was more conservative, with relatively less ***area mapped*** being classified into the moderate and high severity classes, relative to the Histogram Matching and Meddens methods which classified majority of the ***area mapped*** into the high severity class.

These differences in severity classification methodologies of trees per acre ADS data collected into percent-based low, moderate, and high severity classes led to substantially divergent ***damage area*** totals across the study areas. These differences are depicted by the degree of down-weighting that modified ***area mapped*** (area inclusive of dead tree damage and adjacent live tree) to ***damage area*** (area only inclusive of dead tree damage). The $FHP_{R1\&R4}$ method effectively down-weighted ***area mapped*** extents by approximately 70% whereas the Histogram-matching and Meddens methods down-weighted 40% (Figure 8).

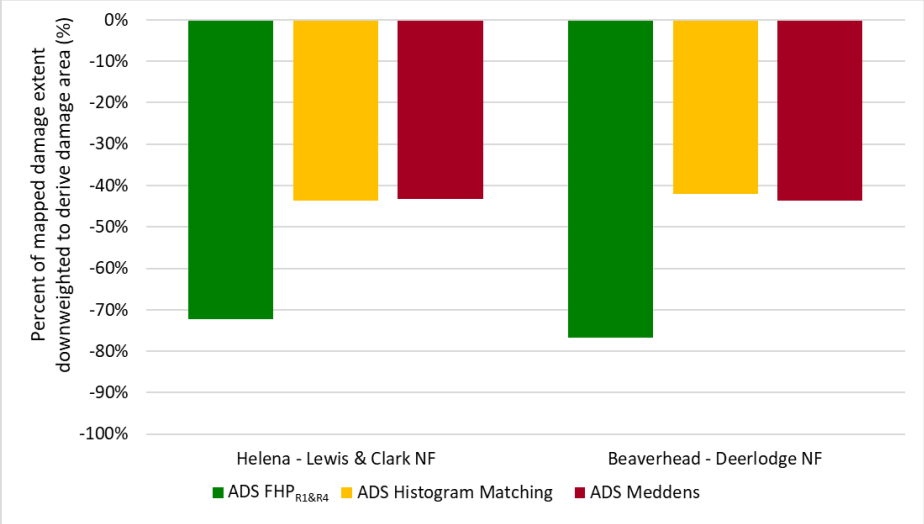

Figure 8: Percent of *area mapped* down-weighted to derive *damage area* on the Helena – Lewis & Clark and Beaverhead – Deerlodge National Forests.

Overall, the $FHP_{R1\&R4}$ method yielded 600k acres of less ***damage area*** relative to Histogram-matching and Meddens methods, with final ***damage area*** totals being very similar to R1-VMap reference data. ***Damage area*** calculated with the $FHP_{R1\&R4}$ method was just under (-15%) to reference data totals for the BVRD NF study area and slightly above (+52%) totals for the H-LC NF. In contrast, the Meddens and Histogram Matching methods far overestimated ***damage area*** totals and yielded 2-3x greater damage area compared to the R1-VMap reference data. These overestimations represented just over 1 million acres of MPB-caused ***damage area*** across each study site.

Discussion

Evaluation of accuracy across multiple scales allows for inferences on which sources of errors can be attributed to ADS data collection and how various post-processing techniques implemented perform at reducing or enhancing the importance of those errors.

Polygon-level Spatial Location, Omission, and Commission Accuracy

The polygon-level analysis of spatial location error indicated ADS damage was mapped in correct locations, and was not mapped in incorrect locations, across 72% of the combined, 11 million-acre study area. In areas ADS was mapped in incorrect locations, omission errors (false negatives) were minimal and represented 7% of the combined study areas while commission errors (false positives) were substantial and represented 22%. Across study areas total ***area mapped*** by ADS was over-estimated 2.7-fold (using $FHP_{R1\&R4}$ post-processing). Visual map interpretation indicated the overestimation was caused by liberal polygon extent mapping as the commission errors were located adjacent to validated damage (see Figures 1 and 2).

Three factors could play a role in the substantial commission errors detected from spatial extents mapped. First, ADS mapping 'spillover' as small, spatially precise polygons are difficult to draw with mapping software limitations and surveyor logistics within high-speed moving aircraft. Second, ADS polygons that

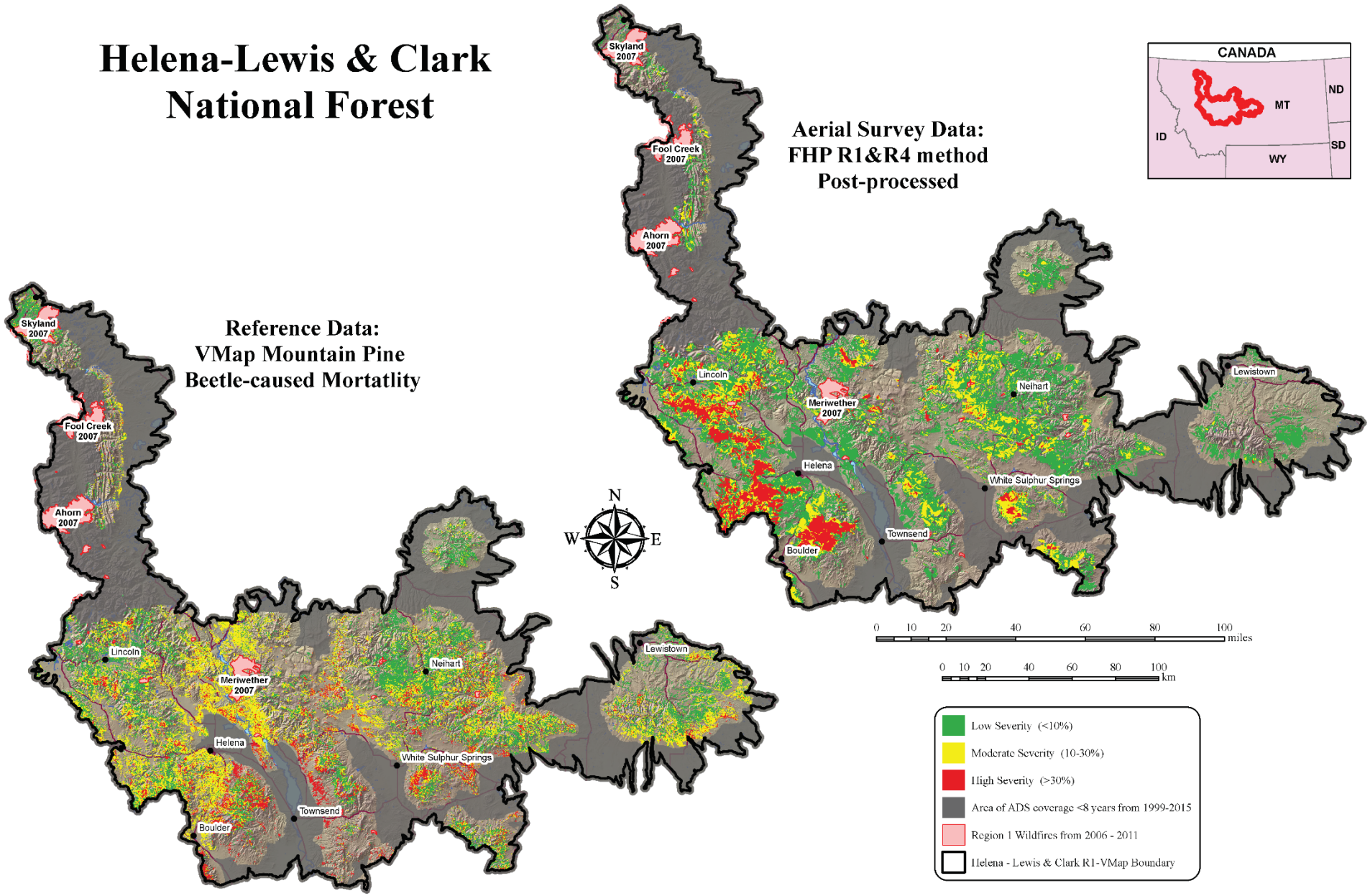

Figure 9a: Mountain pine beetle-caused mortality from 1999-2015 for R1-VMap reference and aerial survey data comparisons on the Helena – Lewis & Clark National Forests.

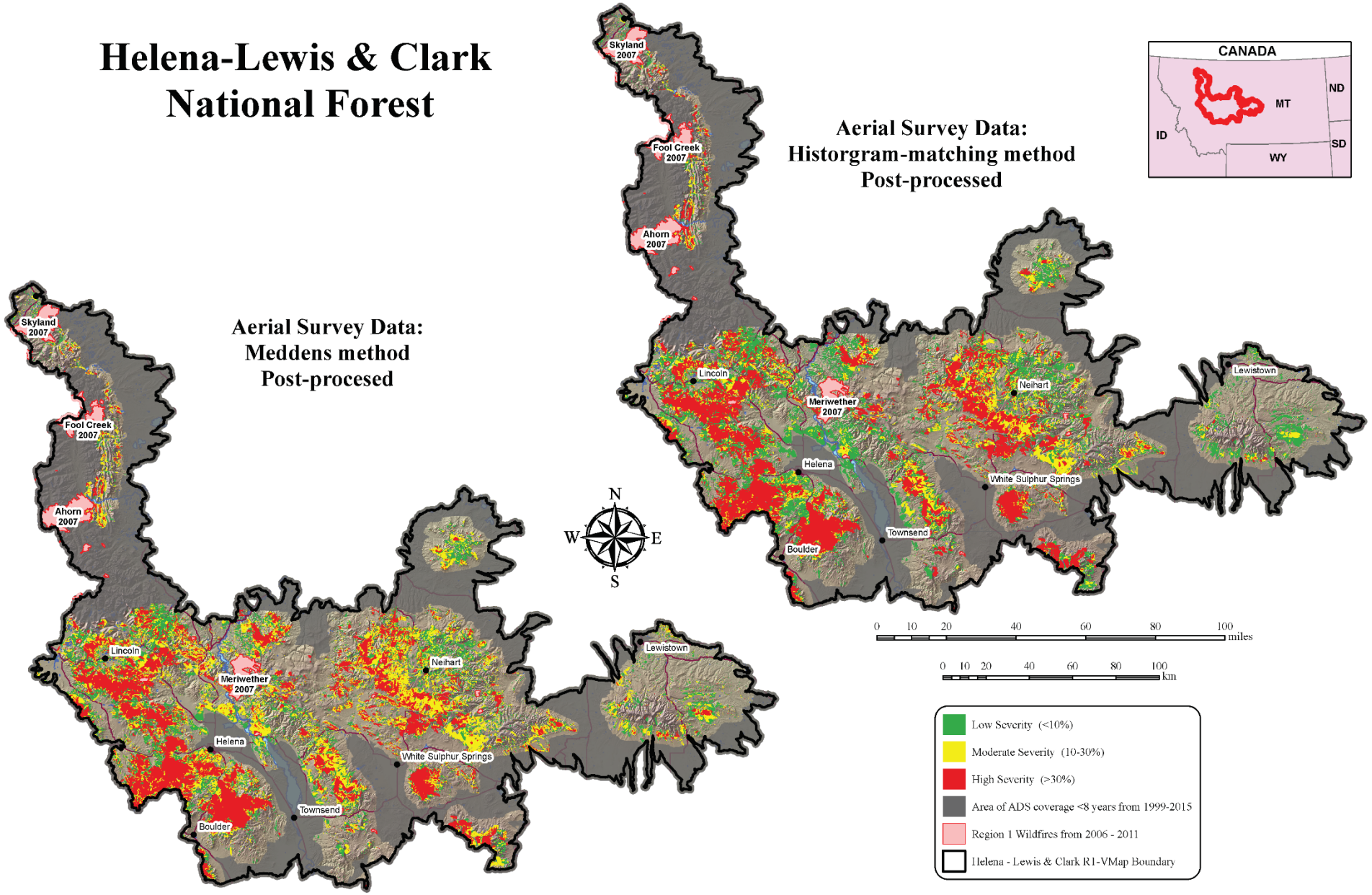

Figure 9b: Mountain pine beetle-caused mortality from 1999-2015 for R1-VMap reference and aerial survey data comparisons on the Helena – Lewis & Clark National Forests.

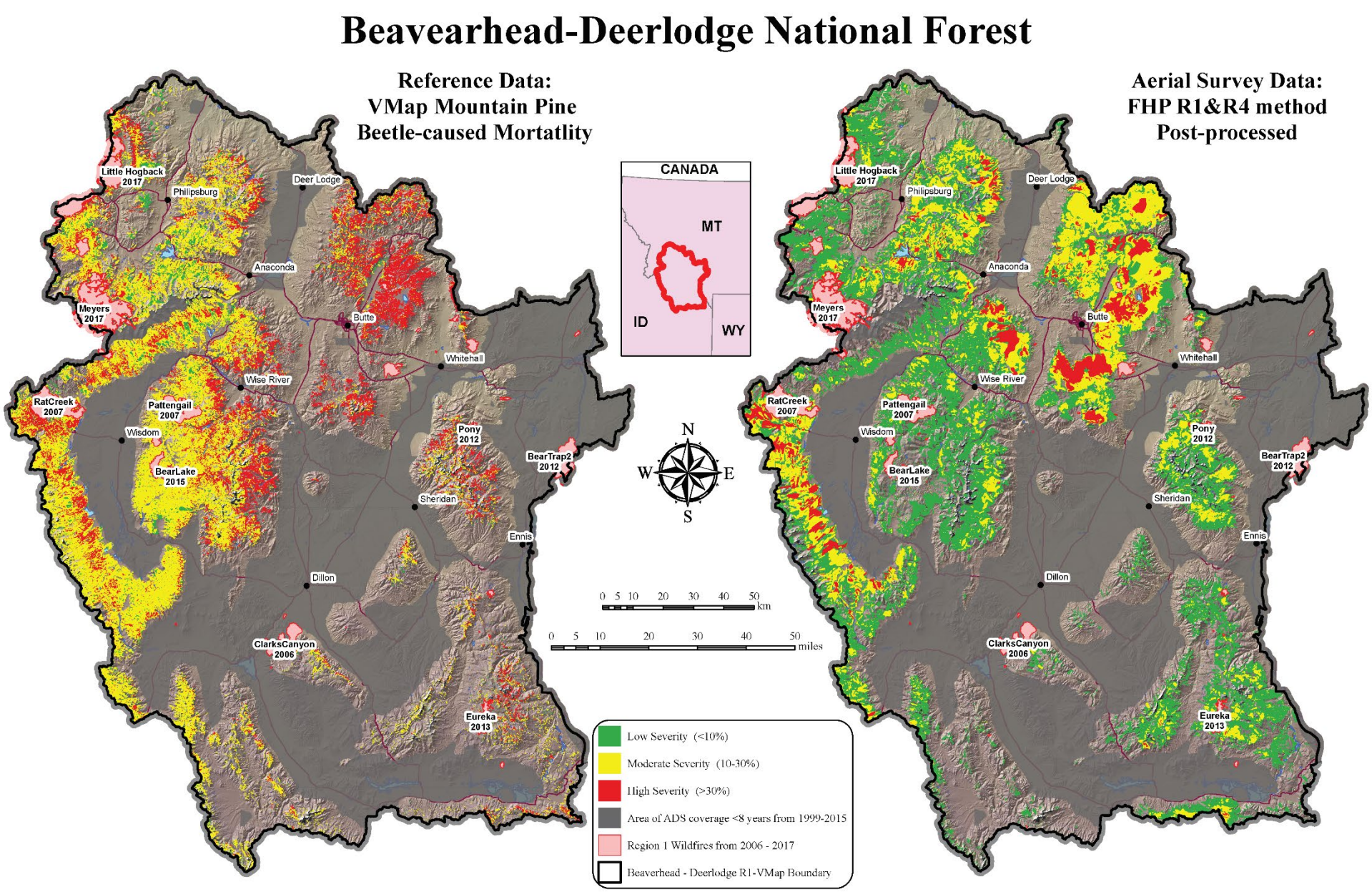

Figure 10a: Mountain pine beetle-caused mortality from 1999-2015 for R1-VMap reference and aerial survey data comparisons on the Beaverhead - Deerlodge National Forests.

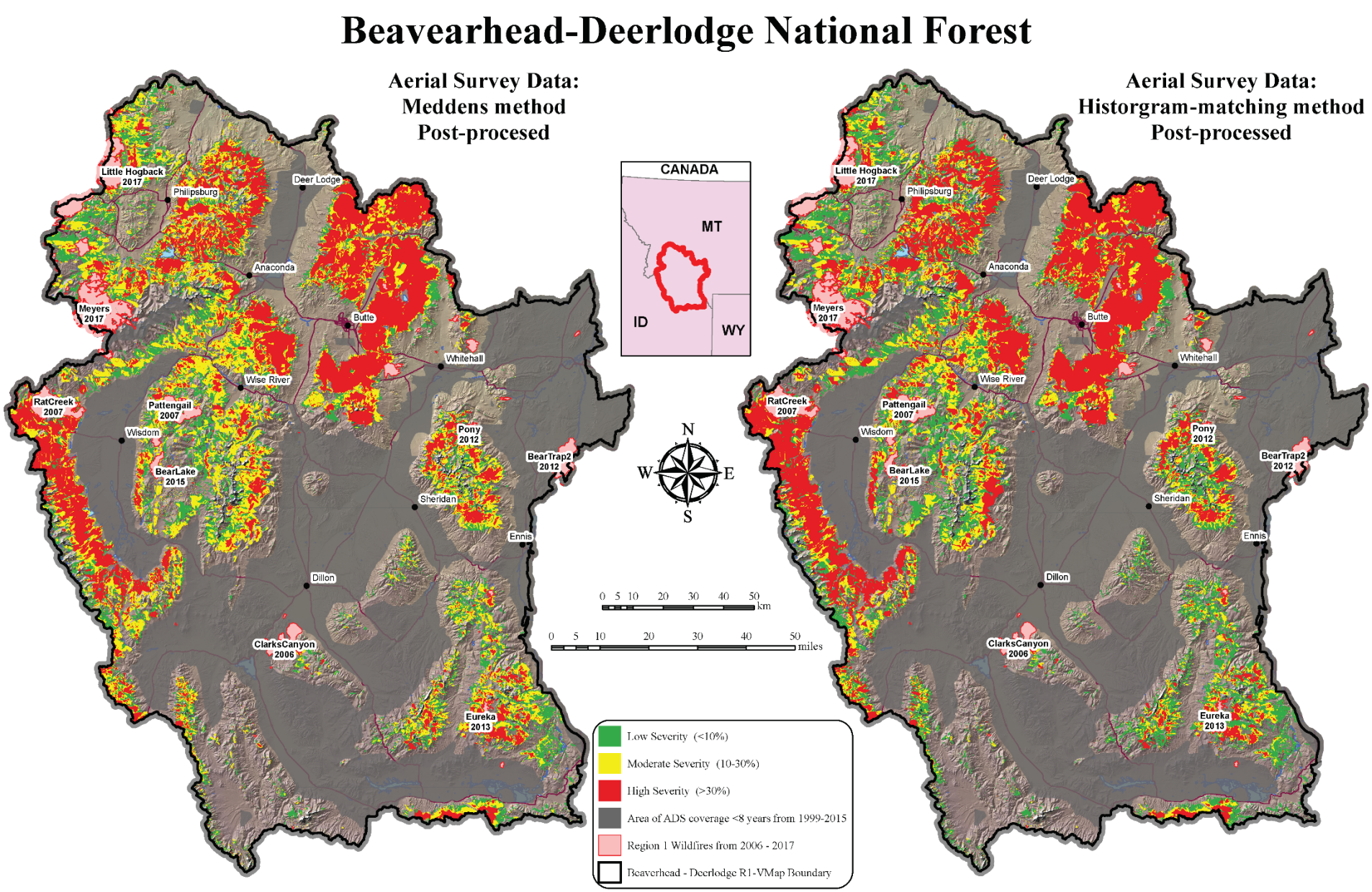

Figure 10b: Mountain pine beetle-caused mortality from 1999-2015 for R1-VMap reference and aerial survey data comparisons on the Beaverhead - Deerlodge National Forest

are aggregated from multiple, yearly surveys spatially combine annual surveyor renditions of mortality locations which may vary slightly each year based on surveyor precision for the exact areas mapped. Third, R1-VMap mortality reference data accuracy, while high, is not 100% for various attributes derived (see Appendix A).

Of these possibilities, the third is unlikely as R1-VMap products utilize NAIP imagery as part of the vegetation mapping process which is considered to have very high spatial location precision. The second option is possible, but considered unlikely as well as surveyors take special care to review prior-year damages mapped prior to and/or during survey missions. The first factor related to difficulties in mapping small-area polygons due to mapping software limitations and survey condition logistics is likely the primary driver of spatial extent overestimations.

Consistent findings of spatial extent overestimation are consistently documented within other ADS accuracy assessment reports. Findings from evaluation of ADS accuracy in the western U.S. reported ADS-estimated spatial extent of mortality was 3.5-fold greater than Worldview II high-resolution reference satellite imagery when ***area mapped*** totals were compared for polygons with beetle-caused mortality (Coleman et al. 2018; see Table 6, Part D).

Similarly, a comprehensive comparison of ADS mapping and aerial photo-interpreted mortality polygons in South Dakota indicated a 4.3-fold increase in total ***area mapped*** by ADS as compared to the aerial photography-based reference data (Backsen and Howell 2013; see Figure 2 and Table 2). Our results of 2.7-fold increase in ADS ***area mapped*** across combined study areas are similar in magnitude to these studies.

Polygon-level Severity Class Accuracy

Severity accuracy was also assessed at the polygon-level, within three percent mortality classes, after post-processing methods converted trees per acre estimates to percent mortality classes for a subset of area where ADS and reference data both indicated damage occurred. ADS-estimates of severity were accurate across 29-47% of the subset area depending on study location and post-processing method used (Table 3). The $FHP_{R1\&R4}$ method had greater proportion of area that underestimated severity while the Histogram-matching and Meddens methods overestimated severity across majority of area.

These results differ slightly from a recent severity assessment conducted in northern Idaho as Bright et al. (2020) identified slightly higher 49-56% classification accuracy with the same post-processing methods utilized for three percent mortality classes. However, it should be noted that accuracy was defined differently as our study calculated percent of total area while Bright et al. (2020) calculated percent of total polygons with correct severity classification. This latter study also found increased overall accuracy (71-78%) when using a ± 15% validation threshold with continuous percent mortality severity estimates for ADS data (calculated from data collected with trees per acre in some years and percent mortality classes in others) and reference imagery (Bright et al. 2020). Similar calculations were not feasible for our ADS dataset as severity was estimated solely with trees per acre metrics.

Aggregation of Spatial Extent and Severity Data for Cross-scale Error Balancing

Spatial extent and severity errors were considered separately to investigate error orientations at the scale in which both ADS-estimates and R1-VMap reference data are attributed. This consideration is critical to understanding how each post-processing method impacted ***damage area*** calculations. These calculations combine both attributes estimated by aerial surveyors to 1) eliminate portions of ADS mapped extents that

represent area comprised of live trees; 2) ensure data are in a comparable scale for reporting and trend analysis; and 3) reduce overall importance of known sources of bias within ADS data (see Egan et al. 2019 for discussion).

The spatial extent errors were substantially oriented towards overestimation, while the severity errors were oriented towards underestimation for the $FHP_{R1\&R4}$ method and overestimation for the Histogram Matching and Meddens methods. Thus, when ***damage area*** calculations occurred, the $FHP_{R1\&R4}$ method down-weighted ***area mapped*** to a far greater degree (approx. 70%) than the Histogram-Matching and Meddens methods (40%, respectively). This occurred as the $FHP_{R1\&R4}$ method was intentionally built to be conservative in its estimations so that total ***damage area*** would not be over-representing actual mortality on the ground when consumed by resource manager and the public end-users (Egan et al. 2019).

Subsequently, as ***damage areas*** were summed to subwatershed units for base-level statistical analysis, the errors of ADS mapped extent overestimation and severity underestimation with the $FHP_{R1\&R4}$ method balanced well and yielded very similar results compared to reference data (+92 acre deviation averaged across subwatersheds). Conversely, the alternative methods had greater high-severity mortality classifications which compounded errors and yielded divergent results relative to reference data (+1400 acre average deviation).

Similar error balancing occurred when data were aggregated from base- to mid-level National Forest ***damage area*** summations. Across subwatersheds, the $FHP_{R1\&R4}$ method also yielded a balanced distribution of subwatersheds that slightly over- and underestimated ***damage area*** while the alternate methods were heavily skewed towards overestimations (see Figure 4 and Table 4). These error distributions were further balanced when aggregated to mid-level for the $FHP_{R1\&R4}$ method and compounded with the alternate methods. At this mid-scale, ***damage area*** was within 0.85 to 1.5-fold, while being 2-fold overestimated with each of the alternate methods (see Figure 7).

Overall, our study identified that the $FHP_{R1\&R4}$ post-processing methodology is preferred as it reduced importance of errors associated with spatial extent overestimation based on its conservative approach to classification of trees per acre to percent severity classes. The Histogram-Matching and Meddens methods yielded results divergent from reference data and substantially overestimated ***damage area*** that depicted impact from the multi-year MPB outbreak studied. These results are consistent with Hicke et al. (2020) which evaluated these methods for utility in bark beetle ***damage area*** calculations at a National-scale and, following logical evaluation, concluded that the $FHP_{R1\&R4}$ method was the most reasonable as it did not substantially over- or underestimate ***damage area***. Our empirical assessment provided in this study verify the Hicke et. al (2020) conclusions and validate the summaries of bark beetle ***damage area*** provided in that study for the western U.S.

Conclusions

This study represents the most comprehensive, multi-scale analysis of ADS data conducted to-date evaluating multiple aspects of potential error within ADS data across an 11 million-acre extent.

Our findings indicate that ADS data, when post-processed appropriately, can have high levels of accuracy to represent multi-year, cumulative bark beetle outbreak impacts. This accuracy was greatest for the location of mortality and ***damage area*** totals, when using the $FHP_{R1\&R4}$ methodology. Aggregating this ***damage area*** to base- and mid-level were found to have cross-scale bias reduction benefits with this method. Thus, we conclude that ADS data, aggregated with post-processing techniques, can provide for accurate forest health renditions.

However, conclusions from this and other studies have indicated how substantial overestimations of bark beetle damages can occur without post-processing. Issues arise when only the ***area mapped*** (also called 'acres with mortality') attribute is the metric for reporting as 1) ***area mapped*** overestimations can be common in ADS data; 2) inclusion of divergent proportions of live and dead tree areas within ***area mapped*** totals are not appropriate to aggregate for reporting summed area totals as they incorporate non-standardized area impacted percentages; and 3) bias from #1 and #2 can compound bias as data are aggregated from polygon-level to coarser scales when errors are unbalanced. Our validation indicates the $FHP_{R1\&R4}$ post-processing methodology effectively corrects for these issues and provide results valid for reporting and analysis at base-level and coarser mid-, Regional-, and National-scales.

Citations

Ahl, R., Gregory, J., Brown, S., David, K., Kaiden, J. and Kellner, F. 2018. The Beaverhead-Deerlodge National Forest Region1 Existing Vegetation Database (VMap) Revision of 2018. Project Report: NRGG_PR_B-D_VMAP2018_09_15_2018, U.S. Department of Agriculture Forest Service, Northern Region Geospatial Group (NRGG), Missoula, MT. 43p.

Backsen, J., Howell, B. 2013. Comparing Aerial Detection and Photo Interpretation for Conducting Forest Health Surveys. Western Journal of Applied Forestry. 28, 3-8.

Berglund, D., Bush, R., Barber, J., and Manning, M. 2009. R1 Multi-level Vegetation Classification, Mapping, Inventory, and Analysis System. Numbered Report 09-01 v2.0, Missoula, MT: U.S. Department of Agriculture Forest Service, Region 1, Forest and Range Management & Engineering, 14p. Accessed July 10, 2019.

Bright, B.C., Hudak, A.T., Egan, J.M., Jorgensen, C.L., Rex, F.E., Hicke, J.A., Meddens, A.J. 2020. Using Satellite Imagery to Evaluate Bark Beetle-Caused Tree Mortality Reported in Aerial Surveys in a Mixed Conifer Forest in Northern Idaho, USA. Forests. 11, 529.

Brohman, R. J., and Bryant, L.D. 2005. Existing Vegetation Classification and Mapping Technical Guide v1.0.Gen. Tech. Rep. WO-67, Washington DC: U.S. Department of Agriculture Forest Service, Ecosystem Management Coordination Staff, 305p. Accessed July 10, 2019.

Coleman, T., Graves, A., Heath, Z., Flowers, R., Hanavan, R. Cluck, D., Ryerson, D. 2018. Accuracy of aerial detection surveys for mapping insect and disease disturbances in the United States. Forest Ecology and Management. 430, 321-336.

Egan, J., Kaiden, J., Lestina, J., Stasey, A. and Jenne, J. 2019. Techniques to Enhance Assessment and Reporting of Pest Damage Estimated with Aerial Detection Surveys. R1-19-09 USDA Forest Service. U.S. Department of Agriculture, Forest Service, Forest Health Protection, Missoula.

European Space Agency. 2015. Sentinel-2 User Handbook. ESA Standard Document, Issue 1, Revision 2, 64p. Accessed July 10, 2019.

Hicke, J.A., Xu, B., Meddens, A., Egan, J. 2020. Characterizing recent bark beetle-caused tree mortality in the western United States from aerial surveys. Forest Ecology and Management. Volume 475.

Johnson, E.W. and Ross, J. 2008. Quantifying error in aerial survey data. Australian Forestry. 71:3, 216-222.

Kaiden, Jeffrey D. 2018. The Region 1 Existing Vegetation Mapping Program (2017). Story Map. U.S. Department of Agriculture Forest Service. Accessed July 17, 2019.

Lestina, J., Egan, J., Harley, G., Jenne, J., Kaiden, J. and Sontag, S. 2019. Broad-level Reconstruction of Mountain Pine Beetle Outbreaks from 1999-2015 across the Northern Region. R1-19-06. USDA Forest Service. U.S. Department of Agriculture, Forest Service, Forest Health Protection, Missoula. 55 p.

Man, G. 2012. Major forest insect and disease conditions in the United States: 2011. FS–1000. USDA Forest Service, Forest Health Protection, Washington, DC. 40 p.

McConnell, T., Johnson, E., Burns, B. 2000. A guide to conducting aerial sketchmapping surveys. FHTET 00-01. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team. 88 p.

Meddens, A., Hicke J., Ferguson, C. 2012. Spatial and temporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western US. Ecol. Appl. 22, 1876–91.

Nelson, M.L.; Brewer, C.K.; Solem, S.J., eds. 2015. Existing vegetation classification, mapping, and inventory technical guide, version 2.0. Gen. Tech. Rep. WO–90. Washington, DC: U.S. Department of Agriculture, Forest Service, Ecosystem Management Coordination Staff. 210 p.

SAS Institute Inc., 2004. SAS/STAT 9.1 User's Guide. SAS Institute Inc., Cary, NC.

U.S. Department of Agriculture Farm Service Agency. 2011. National Agriculture Imagery Program (NAIP) Information Sheet.Accessed July 10, 2019

U.S. Department of Agriculture Forest Service. 2016. Fire Statistics System (FIRESTAT) User Guide for the Individual Wildland Fire Report, FS-5100-29, Kansas City, MO: National Fire and Aviation Management Web Database, National Information Technology Center (NITC), 87p. Accessed July 12, 2019.

U.S. Department of Agriculture Forest Service. 2018. Geospatial Data: Northern Region Existing Vegetation Mapping Program Accessed July 10, 2019.

U.S. Geological Survey and U.S. Department of Agriculture, Natural Resources Conservation Service. 2013. Federal Standards and Procedures for the National Watershed Boundary Dataset (WBD)(4 ed.). U.S. Geological Survey Techniques and Methods 11-A3, 63p. Accessed July 10, 2019.

Zweifler, M. 2019. Digital Mobile Sketch Mapping (DMSM): Procedure for Cross walking TPA and Percent Affected. USDA Forest Service. pg. 2.

Citation for this report:

Egan, J., Kaiden, J. and Lestina, J. 2020. Multi-Scale Validation of Aerial Detection Survey Accuracy and Post-Processing Techniques to Depict Mountain Pine Beetle Outbreak Severity and Extent in the Northern Region. R1-20-05 USDA Forest Service. U.S. Department of Agriculture, Forest Service, Forest Health Protection, Missoula MT. 25 p.